JOURNAL OF MASS SPECTROMETRY
J. Mass Spectrom. 2005; **40**: 1472–1483
Published online 1 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jms.938

JMS

Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography – electrospray mass spectrometry

Tímea Imre,[1] Gitta Schlosser,[2] Gabriella Pocsfalvi,[3] Rosa Siciliano,[3] Éva Molnár-Szöllősi,[4] Tibor Kremmer,[4] Antonio Malorni[3] and Károly Vékey[1]*

[1] Department of Mass Spectrometry, Institute of Structural Chemistry, Chemical Research Center, Hungarian Academy of Sciences, 1525 Budapest, Hungary
[2] Research Group of Peptide Chemistry, Hungarian Academy of Sciences, Eotvos L. University, Budapest 112, Hungary
[3] Instituto di Scienze dell' Alimentazione, Consiglio Nazionale Delle Ricerche, Avellino, Via Roma 52 a-c. 83100, Italy
[4] Department of Biochemistry, National Institute of Oncology, H-1122 Budapest, Hungary

Received 13 May 2005; Accepted 14 September 2005

A new anionic surfactant (RapiGest SF) was successfully used for site-specific analysis of glycosylation in human alpha-1-acid glycoprotein (AGP). By means of this analytical approach combined with capillary HPLC-mass spectrometry (and tandem mass spectrometry), the N-linked glycosylation pattern of AGP was explored. On the basis of mass matching and MS/MS experiments *ca* 80 different AGP-derived glycopeptides were identified. Glycosylation shows a markedly different pattern for the various glycosylation sites. At sites I and II, triantennary complex-type oligosaccharides predominate and at sites III, IV and V, tetra-antennary complex-type oligosaccharides predominate. Sites IV and V show the presence of additional *N*-acetyl lactosamine (Gal-GlcNAc) units (even higher degree of branching and/or longer antennae are also present).

KEYWORDS: alpha-1-acid glycoprotein; RapiGest SF; glycosylation; glycopeptide; mass spectrometry

INTRODUCTION

One of the common ways of posttranslational modification is the process of glycosylation, in which oligosaccharides are attached to specific sites encoded in the primary sequence of the proteins.[1] Glycosylation plays a varied role in the structure and function of proteins,[2] and changes in glycosylation can predict numerous diseases including cancer and immune system deficiencies.[3,4] Oligosaccharides can influence protein folding, affect solubility of proteins, sterically hinder protease digestion and act as a recognition signal as well. The biological synthesis and processing of glycoprotein carbohydrates result in microheterogeneity owing to the heterogeneous populations of structurally related oligosaccharides. In addition, various glycosylation sites are substituted by different populations of carbohydrates usually described as site heterogeneity.[5] Determination of location, heterogeneity and structure of carbohydrate moieties is therefore important in understanding the biological functions of glycoproteins.

Mass spectrometry is among the best-suited techniques that can provide structural information on glycoproteins.[6] Glycoproteins are usually characterized by two common techniques. First, glycans are released from the glycoproteins enzymatically or chemically, and following purification, the oligosaccharides are analyzed by mass spectrometry.[7–10] Although it results in simpler mass spectra, information about site occupancy is lost. The other approach is to use enzymatic digestion of glycoproteins with endoproteases and characterize the resulted glycopeptides by HPLC–MS, which provides site analysis of glycosylation.[11–13] Proteolysis is usually achieved with trypsin, GluC or other nonspecific proteases,[13,14] and the peptide–glycopeptide mixture is then analyzed by on-line reversed-phase high-performance liquid chromatography–electrospray ionization mass spectrometry (RP-HPLC–ESI-MS). In mixtures, glycopeptides usually have lower desorption/ionization efficiency than peptides, making glycopeptide analysis more difficult. Furthermore, glycopeptides are often eluted very early and are masked by 'junk' eluted at the beginning of the separation/purification. In the analysis of glycosylation, the mass spectrometer can be used as a selective detector for glycopeptides, by monitoring of specific and diagnostic sugar oxonium ion fragments[11,12] (selected ion monitoring of ions m/z 204: HexNAc, 274: NeuAc-H_2O, 292: NeuAc,

*Correspondence to: Károly Vékey, Chemical Research Center, Hungarian Academy of Sciences, H-1525 Budapest, Hungary.
E-mail: vekey@chemres.hu
Contract/grant sponsor: Center of Excellence Project (CBCH); Contract/grant number: QLK2-CT-2002-90436.
Contract/grant sponsor: Education Ministry of Hungary NKFP MediChem2; Contract/grant number: 1/A/005/2004.
Contract/grant sponsor: Italian-Hungarian Joint Project 2004/2006; Contract/grant number: MTA707.

366: Hex-HexNAc, 657: NeuAc-Hex-HexNAc) produced in the source/interface region of an electrospray ion source if high orifice (declustering) potential is used. As an alternative, precursor ion scanning in triple quadrupole–type instruments may also be used. This detects the parent ions in the first quadrupole (Q1) that give rise to the characteristic fragment ions created by collisions in the second quadrupole (Q2).

It is known that enzymatic digestion of insoluble or proteolytically resistant proteins is difficult and generates only limited amounts of peptides.[15–17] High peptide coverage is especially desirable when protein modifications, especially posttranslational modifications, are studied. Large carbohydrate moieties in glycoproteins do hinder protease activity; hence, trypsin or other proteases often miss cleavage sites close to a glycosylated asparagine. To improve the efficiency of digestion, denaturating agents such as surfactants (urea; guanidine hydrochloride, GHCl; sodium dodecylsulfate, SDS and organic solvents) are often employed. These reagents improve the unfolding and solubility of proteins, but often cause undesirable chemical modifications and also reduce the efficiency of proteolytic digestion. These reagents, however, often interfere with liquid chromatography and mass spectrometry and require complicated sample cleanup (dialysis, solid-phase extraction, reversed-phase or hydrophobic interaction chromatography, surfactant precipitation) prior to analysis. A new acid-labile surfactant (sodium-3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate) under the trade name of RapiGest SF (Waters, Millford, MA) has recently been introduced for in-solution digestion.[18] It has been designed to eliminate adverse side effects of the conventional denaturants. RapiGest SF effectively unfolds and solubilizes proteins and makes them amenable to proteolysis. It does not inhibit proteolytic activity of trypsin, GluC, LysC, AspN and chymotrypsin.[18] Furthermore, owing to its acid-labile character, sample preparation prior to LC–MS or MALDI-MS analysis consists only of simple acidification. Addition of a strong acid (HCl, TFA) after digestion quickly degrades RapiGest into products that do not interfere with mass spectrometric analysis. Applying this new surfactant, the digestion of various proteolysis-resistant globular and membrane proteins has been achieved.[19,20] Application of the RapiGest for digestion of a monoclonal antibody was found advantageous compared to classical GHCl denaturant.[18] In the present study, efficiency of this sample preparation method has been tried for glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) combined with LC–MS and MALDI-MS analysis.

Human AGP also called *orosomucoid* (ORM) is an acute-phase plasma protein with 41–43 kDa of molecular weight, containing 45% carbohydrate content attached in the form of five N-linked complex glycans.[21,22] The peptide moiety is a single chain of 183 amino acids[23] encoded by three different genes causing genetic polymorphism. The AGP A gene encodes the major variants ORM1 (ORM1∗F1, ORM1∗F2 and ORM1∗S alleles) and the AGP B/B′ genes the variant ORM2 (ORM2∗A). AGP B/B′ genes are identical; the AGP A gene is structurally similar to them, but contains 22 codon/base substitutions.[23] ORM1 variants F1, F2 and S are encoded by the alleles of the same gene and differ in less than five amino acids (ORM1 F1 has Gln-38/Val-174; ORM1 F2 has Gln-38/Met-174 and ORM1 S has Arg-38/Val-174). ORM2∗A variant is the product of the B/B′ gene and exists in several different variants (based on the SwissProt database in 2005, http://us.expasy.org/cgi-bin/niceprot.pl?P02763). The carbohydrate moiety of AGP expressing bi-, tri- and tetra-antennary complex–type *N*-glycans has been investigated extensively.[22] Each of the N-glycosylation sites of AGP (Asn-15, 38, 54, 75 and 85) is occupied by so-called complex-type (variously branched, sialylated and possibly fucosylated) N-linked glycans, resulting in a high degree of heterogeneity. The heterogeneity of AGP glycoforms in plasma may also depend on the pathophysiological conditions of a person.[23] AGP is one of the major positive acute-phase proteins, and although it has been studied over many years, its exact physiological function is still unknown. Substantial increase in glycoforms expressing bi-antennary glycans is noticeable in the early phase of an acute-phase reaction as well as an increase in the degree of fucosylation. Changes in AGP glycosylation in other pathological conditions like chronic inflammation, pregnancy, rheumatoid arthritis, alcoholic liver cirrhosis, sepsis,[21,23–25] just as in cancer,[24–26] are also known. Information about the site occupancy of glycans in AGP is far from being well characterized yet. Treuheit and coworkers studied distribution of oligosaccharides at the five glycosylation sites in the distinct AGP variants by concanavalin A affinity-chromatography, RP-HPLC separation and off-line MALDI-MS analysis. It has been found that the two gene products of ORM are differently glycosylated, and the distribution of bi-, tri- and tetra-antennary chains are different at each site in the three molecular variants.[27] The complexity of the overlapping HPLC peaks was decreased by desialyzation of glycopeptide's fractions. Juhasz and coworkers applied nonspecific endoprotease (pronase) in characterization of glycoproteins and examined the products by MALDI-MS in negative ionization mode.[13] Studying the highly complex AGP with this method, all five glycosylation sites were observed and 27 glycopeptides have been identified. In another paper, Dage studied the site localization of sialyl Lewis^x antigen (NeuAc[alpha]2-3Gal[beta]1–4(Fuc[alpha]1–3)GlcNAc) in AGP by HPLC–MS.[14] A tandem mass spectrometric experiment was developed to detect the presence of sialyl Lewis^x antigen on the basis of the selective detection of m/z 803 sugar oxonium ion produced in the intermediate pressure region of the electrospray interface. The results showed that each of the five N-glycosylation sites contain at least one sialyl Lewis^x antigen which has been implicated as a necessary structure for selectin-mediated cell adhesion during the inflammatory response.[14]

AGP glycan structures have been investigated by our group recently;[25] our present aim was to study glycosylation site occupancy. As AGP proved resistant to conventional tryptic digestion, application of a new surfactant, RapiGest SF, was attempted.

EXPERIMENTAL

Materials

AGP, ammonium bicarbonate (NH_4HCO_3), acetonitrile, trifluoroacetic acid and sequencing-grade trypsin were purchased from Sigma (Hungary). RapiGest SF (lyophilized sodium-3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate) was obtained from Waters (Hungary). Iodoacetamide and dithiothreitol (DTT) were also purchased from Sigma-Aldrich (Hungary). All water used was deionized and purified with a Millipore MilliQ water-purification system.

Enzymatic digestion

Fifty micrograms of AGP was solubilized with 25 µl 0.2% (w/v) RapiGest SF solution, buffered with 50 mM ammonium bicarbonate (NH_4HCO_3). Three microliters of 45 mM DTT in 100 mM NH_4HCO_3 were added and kept at 60 °C for 30 min. After cooling the sample to room temperature, 5 µl of 100 mM iodoacetamide in 100 mM NH_4HCO_3 were added and placed in the dark for 30 min. The reduced and alkylated AGP was then digested by 5 µl (1 mg/ml) trypsin (the enzyme-to-protein ratio was 1 : 10). The sample was incubated at 37 °C for 60 min.

Sample preparation for mass spectrometry

To degrade the surfactant, 4.2 µl of HCl (500 mM) solution was added to the digested AGP sample to have the final concentration between 30 and 50 mM (pH ≈ 2) and was incubated at 37 °C for 45 min. For LC–MS analysis, the acid-treated sample was centrifuged for 10 min at 13 000 rpm. This solution was diluted ten times just before MS analysis.

MALDI-TOF mass spectrometry conditions

MALDI mass spectra were acquired on a Voyager DE-Pro MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA), equipped with a 337 nm nitrogen laser, linear acceleration positive ion mode. Spectra were acquired by 20 kV acceleration voltage, 95% grid voltage, 0.05% guide wire and 180 ns delay time in the range of m/z 500–8000. Mass-to-charge (m/z) calibration was performed externally using a set of peptides as a standard (Perseptive Biosystems). The RapiGest SF treated and trypsin-digested AGP solution (1 µl, 1 pmol) was mixed on the side-inlet probe with matrix (2,5-DHB) dissolved in acetonitrile in a concentration of 10 mg/ml (v/v %) and allowed to dry in air.

Reversed-phase capillary HPLC–MS and MS/MS analysis

The products of the AGP digest were separated on a Waters CapLC HPLC system, equipped with low-flow capillary HPLC pumps and with an autosampler. The mobile-phase A was 0.1% TFA/2% ACN/97.9% H_2O; mobile-phase B was 0.1% TFA/5% H_2O/94.9% ACN. 1.4 µl sample (4 pmol) was injected in microliter pickup mode onto a C18 precolumn (5 × 0.3 mm) and on-line separation was carried out on a C18 (PepMap, 150 × 0.3 mm, 3 µm) column (LCPackings, Amsterdam) with a flow rate of 5 µl/min at room temperature. The initial conditions for separation were 5% B for 3 min, followed by a linear gradient to 20% B by 7 min, 40% B by 50 min then to 100% B by 4 min. Electrospray ionization mass spectrometric experiments were performed on a Micromass Q-Tof Micro (Manchester, UK) hybrid, quadrupole orthogonal acceleration time-of-flight mass spectrometer, interfaced with an orthogonal Z-spray source. The instrument was controlled by Micromass MassLynx 4.0 software and operated in positive ion mode, in the m/z region 100–2200. Desolvation and source temperatures were 150 and 120 °C respectively. Operating conditions optimized for the detection of glycopeptides were the following: capillary voltage 3200 V, sample cone voltage 20 V, extraction voltage 1 V, collision cell voltage 5 V. MS/MS data were obtained in survey scan mode, using 25 V collision energy. MS/MS spectra were recorded in the 100–2200 m/z region and the precursor ion selection was limited to the 1000–2200 m/z region.

RESULTS AND DISCUSSION

Sample preparation

RapiGest SF proved to be very useful for treatment of the usually digestion-resistant AGP. We believe that this is the first time RapiGest SF is used for digestion of a heavily glycosylated glycoprotein, and it did simplify sample treatment significantly.

As an initial step, human serum AGP was treated with the new surfactant before tryptic digestion. RapiGest SF does not interfere with trypsin activity in wide concentration range, allowing rapid in-solution digestion.[19] The time necessary for reduction/alkylation and tryptic digestion is reduced to 4 h against the conventional one-and-a-half-day (and still not completely efficient) procedure. After tryptic hydrolysis, 0.5 M HCl was added to degrade the surfactant, generating product *dodeca-2-one* and *sodium-3-(2,3-dihydroxy-propoxy)propanesulfonate*.[19] After incubation for 45 min at 37 °C, complete degradation of the surfactant was obtained. Note that it is necessary to remove completely the intact RapiGest SF from the sample, because it can act as a strong ion-pairing agent in RP-HPLC conditions and can shift retention times of peptides. Product *dodeca-2-one* is removed by centrifugation, and the other degradation product, *sodium-3-(2,3-dihydroxy-propoxy)propanesulfonate* elutes at the void volume;[19] so, neither interferes with MS analysis.

MALDI-TOF analysis

To check the results of digestion, the sample was subjected to MALDI-MS analysis (positive ion mode). In the MALDI-MS spectrum (Fig. 1), peaks in the lower m/z region correspond to AGP tryptic peptides and in the higher m/z region, correspond to the AGP-derived glycopeptides. No peaks can be detected in the region of m/z 30–40 kDa, indicating that complete digestion of AGP was achieved. AGP-derived glycopeptide peaks in the mass region of 4–6.5 kDa appear with relatively high intensity, while in the low mass region, the most abundant peaks are due to AGP-derived tryptic peptides, providing 65% sequence coverage (55% without RapiGest). Note that although glycopeptides are easily identified in MALDI, their amount is much lower than those of tryptic peptides. Note also that glycopeptide-related

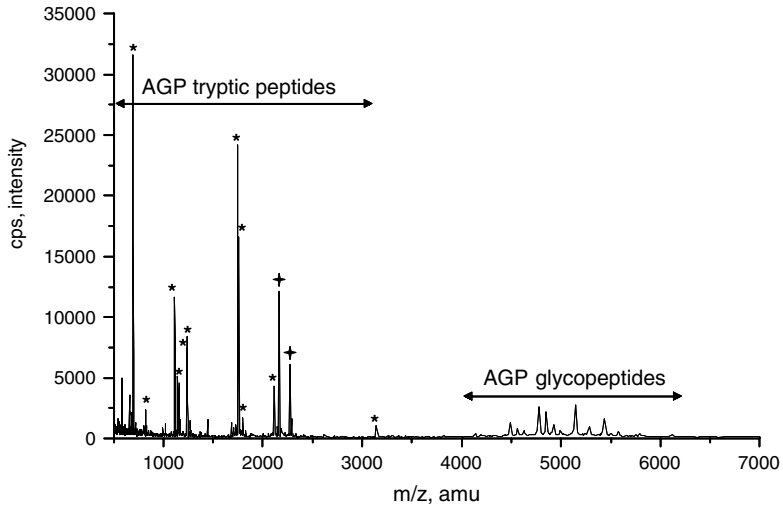

Figure 1. MALDI-TOF/MS spectrum of the AGP tryptic digest, treated with RapiGest SF. * Identified AGP tryptic peptides, + trypsin autodigest products.

peaks in Fig. 1 are quite broad, which is an indication for their heterogeneity. The low resolution of the instrument in the linear mode at higher m/z range is also a reason for the observation. Detailed analysis of glycopeptides and their heterogeneity was performed using HPLC–MS and MS/MS analysis.

HPLC–MS and HPLC–MS/MS analysis

The RapiGest-treated, reduced, alkylated and digested AGP samples were subjected to reverse-phase HPLC directly coupled to an ESI-QTOF instrument. Owing to the high-resolution capability of the instrument, charge states of multiply charged glycopeptide ions could easily be determined. Identifying the structure of a given glycopeptide requires a fairly complex nomenclature. In the present paper the nomenclature of Dage *et al.*[14] is used. Designation bi, tri, tetra, penta, hexa refer to the number of antenna present on the glycan (more accurately, the number of *N*-acetyl lactosamine: Gal-GlcNAc units); S and the following numeral mark the number of sialic acids, while F represents the number of fucose units present. The roman numeral in the front indicates the glycosylation site, while its lower index indicates the genetic variant. For example, the abbreviation V_1-TetraS3F1 marks a glycopeptide at glycosylation site V (Asn^{85}) with peptide sequence derived from ORM1 gene including amino acid sequence [84–90] to which a tetra-antennary trisialylated glycan with one fucose is linked. The structures indicated in the present paper are based on mass matching and confirmed by tandem mass spectrometric analysis.

A typical HPLC chromatogram is shown in Fig. 2. The curve in the top shows the base-peak ion (BPI) chromatogram, which plots the highest intensity ion at each scan. BPI is analogous to the conventionally used total ion chromatogram (TIC), but its use is much preferred for

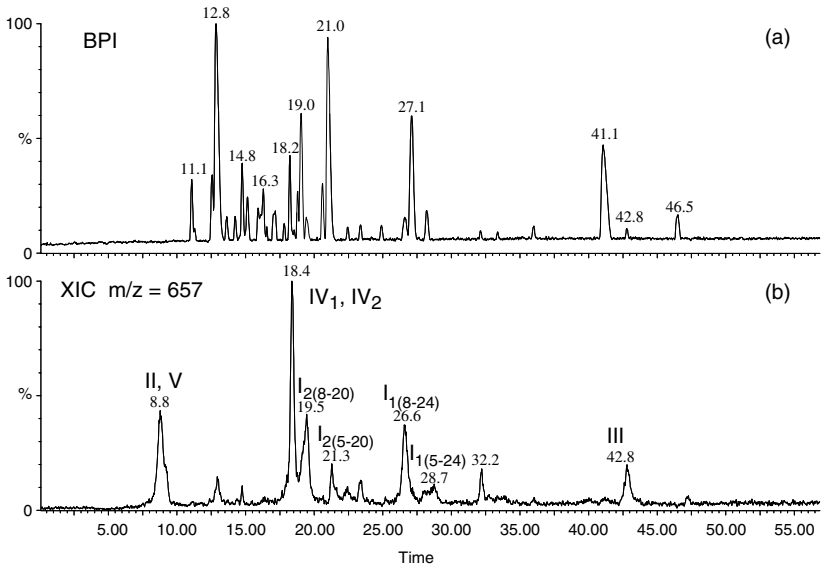

Figure 2. (a) Base-peak chromatogram (BPI) of the AGP tryptic digest and (b) extracted ion chromatogram (XIC) of the diagnostic carbohydrate ion *m/z* 657 (NeuAc-GlcNAc-Gal), which displays the glycopeptide-containing peaks.

complex mixtures (like for protein digests) as it yields a 'cleaner' chromatogram.

The first step of HPLC–MS analysis of glycopeptides was their identification among much more abundant tryptic peptides. (Note that tryptic digestion of AGP yields about five times more peptides than glycopeptides; the latter are heterogeneous owing to the variable sugar content and, furthermore, glycopeptides have a much lower sensitivity than peptides in ESI.) Typical fragmentation of glycopeptides (at high declustering or collision energy) involve formation of diagnostic sugar fragments, like m/z 204 for GlcNAc, m/z 292 and 274 for NeuAc (sialic acid), m/z 366 for Gal-GlcNAc and m/z 657 for NeuAc-Gal-GlcNAc. These ions are often used for identifying glycopeptide-containing chromatographic peaks.[11,12] The lower part of Fig. 2 shows the ion chromatogram of such a diagnostic fragment at m/z 657. Chromatograms of other diagnostic sugar fragments (m/z 274, 292, 366) were very similar (data not shown). On the basis of these diagnostic peaks, it was easy to identify the glycopeptides. Figure 2 also shows that (as expected) the glycopeptides gave only minor peaks in the AGP digest mixture. Note also that glycopeptide-containing chromatographic peaks are much broader than the BPI peaks of peptides. The mass spectra obtained from these chromatographic peaks indicate a series of glycopeptide molecular masses. A typical example is shown in Fig. 3 obtained from the peak at retention time 42.8 min. The results show that retention time is mainly determined by the peptide backbone and those corresponding to different glycosylation vary only very slightly. This is illustrated in Fig. 4, which clearly shows that peak broadening is due to heterogeneity of the oligosaccharide chains attached to the same peptide. This also proves that the various masses observed in the 1000–2000-kDa range in Fig. 3 are indeed molecular ions and not fragments.

Deconvoluting the various charge states in Fig. 3, one obtains a 'cleaner' mass spectrum representative of molecular mass distribution (Fig. 5). The various peaks correspond to a series of glycoforms attached to the glycosylation site III. Note that the corresponding peptide sequence [46–63] contains a missed cleavage site (at the Lys^{55} residue), located just after the glycosylated asparagine (Asn^{54}). It is very likely that the cleavage of trypsin at this site is blocked by the neighboring bulky oligosaccharide. Such complications, although often encountered, complicate glycoprotein analysis. Luckily, the peptide sequence can be (and was) identified by mass matching and by tandem mass spectrometry.

To confirm the suggested glycopeptide structures, MS/MS measurements were also performed. As an example, the MS/MS spectrum of the glycopeptide, assigned as the IV-TetraS3 ion at m/z 1286.6 (4+) at retention time 18.4 min, is shown in Fig. 6. When a glycopeptide undergoes collision-induced dissociation, a characteristic product ion is formed by cleaving the bond between the first and the second GlcNAc unit adjacent to the peptide backbone.[28,29] To identify this peptide, the molecular mass of the GlcNAc residue (203) has to be subtracted from the mass of this ion, and the remainder can be searched for peptide identification. Although there are various other cleavages, the fragmentation process always stops at the peptide + GlcNAc residue and the mass of this unit is easy to identify. In Fig. 6 in the upper full-scan MS/MS spectrum, the mono-, di- and trisaccharide diagnostic sugar-derived oxonium ions corresponding to GlcNAc (204), NeuAc-H_2O (274), NeuAc (292),

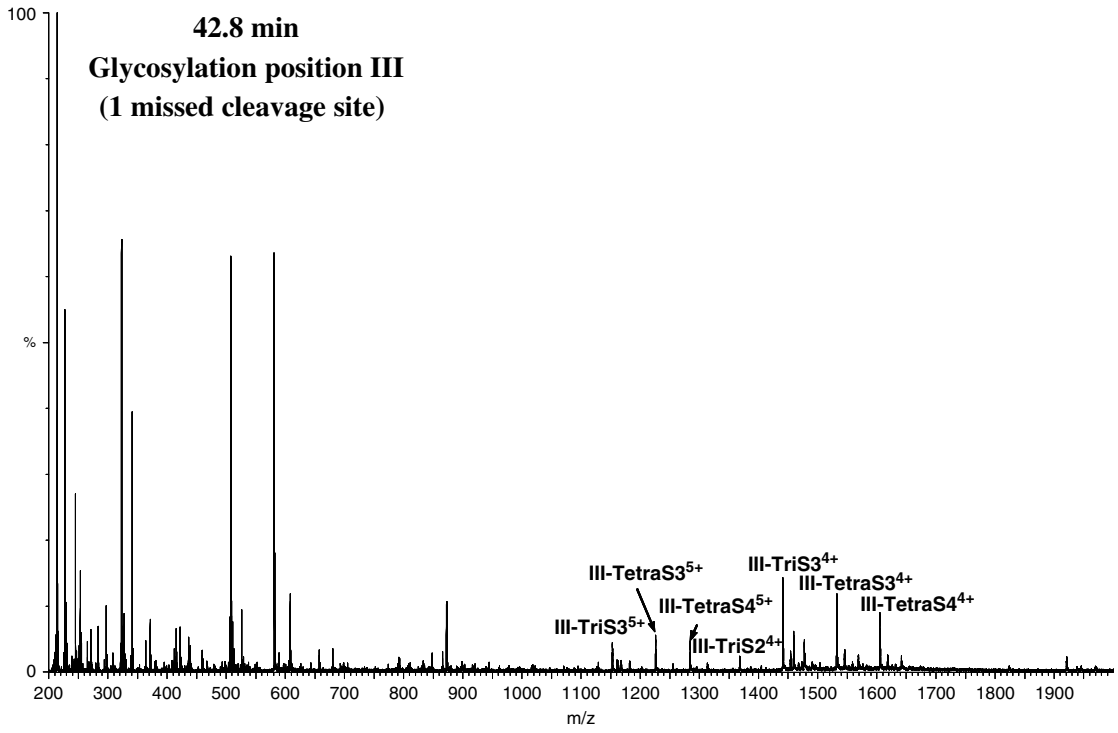

Figure 3. ESI-MS spectrum of a chromatographic peak identified as site III glycopeptide at retention time 42.8 min. The glycopeptide sequences were identified by mass matching and subsequent MS/MS measurements.

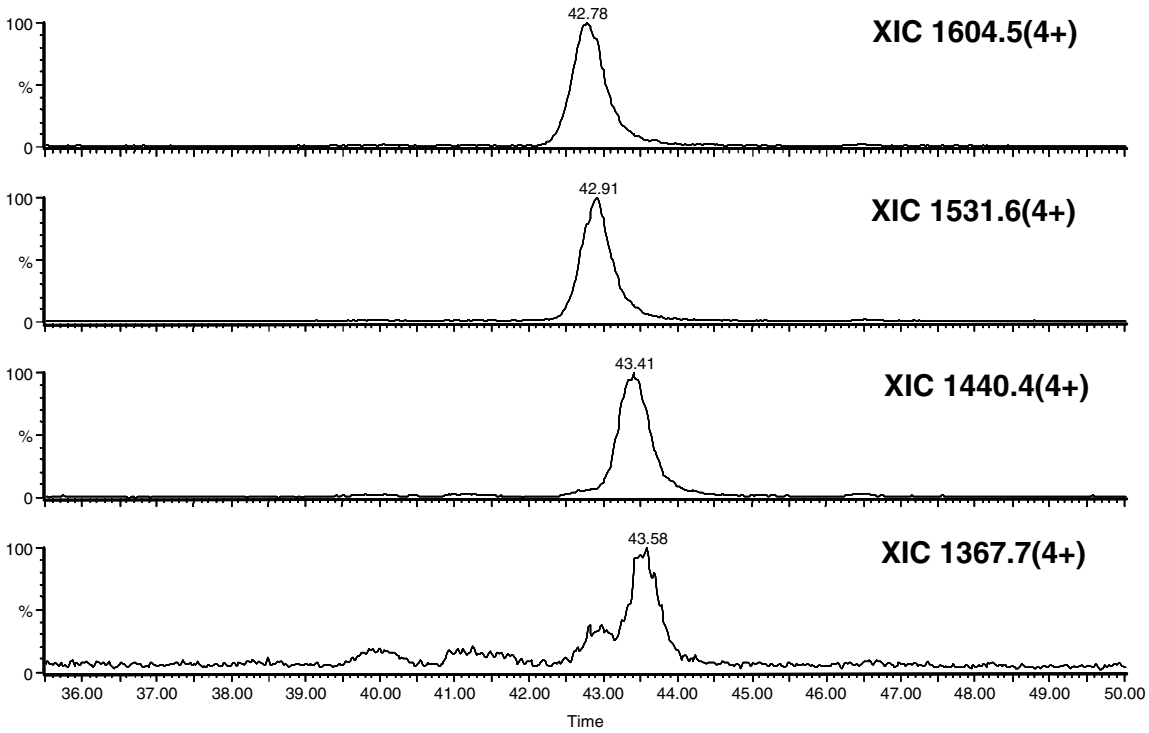

Figure 4. Extracted ion chromatograms (XIC) of ions *m/z* 1604.5 (4+), 1531.6 (4+), 1440.4 (4+) and 1367.7 (4+) corresponding to structures III-TetraS4, III-TetraS3, III-TriS3 and III-TriS2 respectively. The peptide part of the glycopeptide series contains a missed cleavage site.

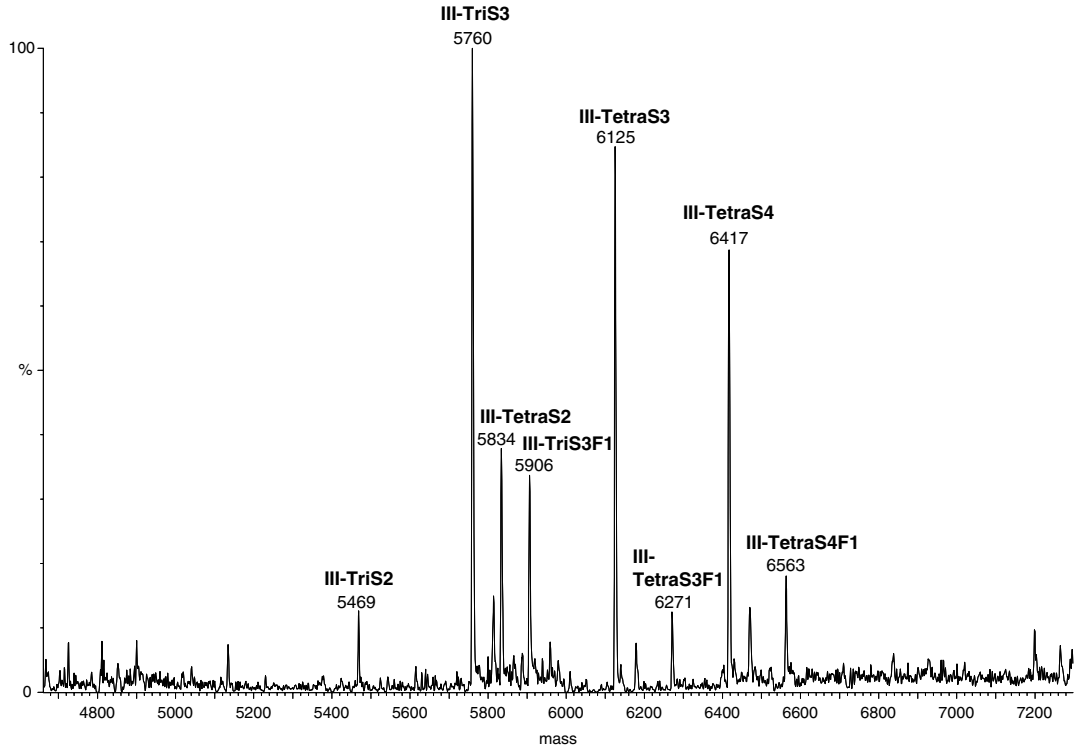

Figure 5. Deconvoluted ES-MS spectrum from a glycopeptide peak identified as site III from HPLC-ESI/MS analysis (42.8 min). The glycopeptide structures were identified by mass matching and MS/MS.

Gal-GlcNAc (366) and NeuAc-Gal-GlcNAc (657) residues dominate in the low mass region, while the larger glycopeptide fragments are observed in the region of *m/z* 1000–2100 Da. The characteristic (peptide + GlcNAc) ion can be found at *m/z* 1060 (2+) with a rather high intensity. The oligosaccharide units usually fragment by loss of NeuAc and NeuAc-Gal-GlcNAc units (corresponding to an antenna with sialic acid), giving information on

JMS

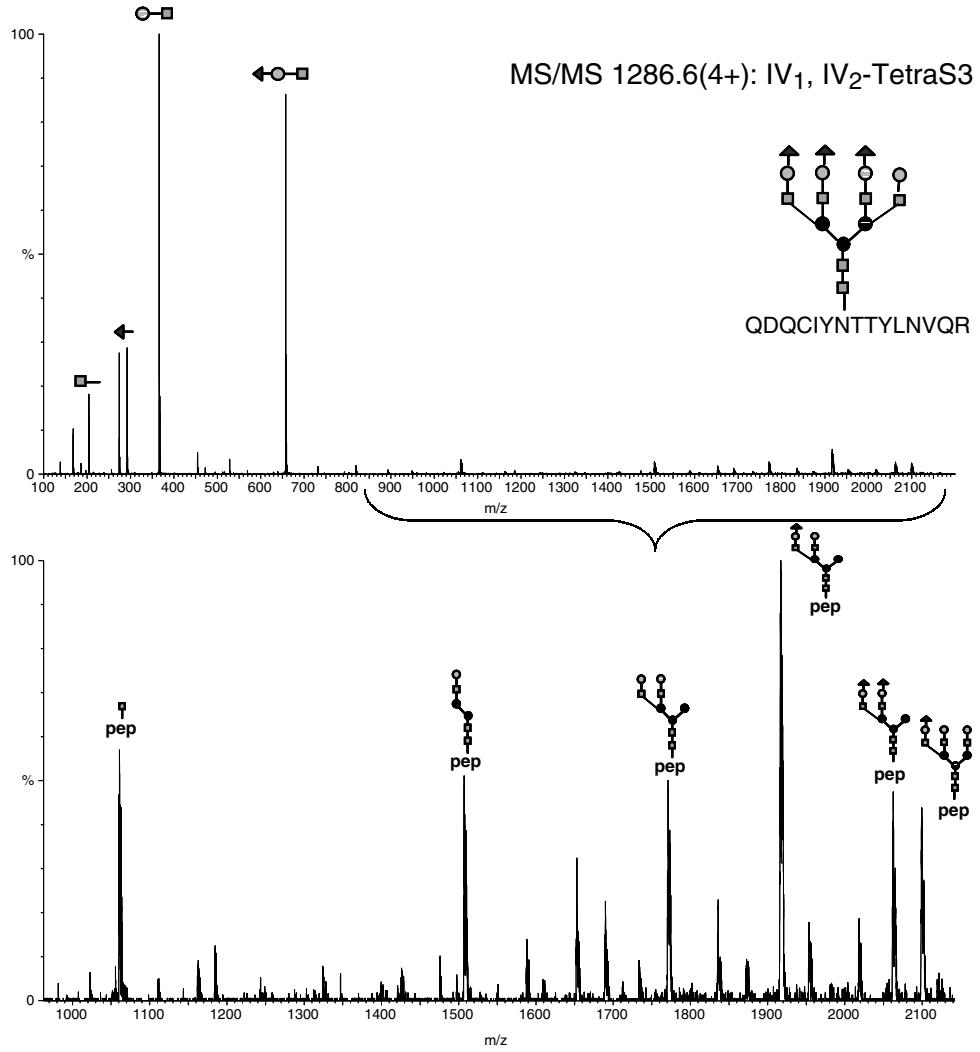

Figure 6. MS/MS spectra of glycopeptide from IV-TetraS3 1286.6 (4+) ion at 18.4 min. ◀ NeuAc, ○ Gal, □ GlcNAc and ● Man.

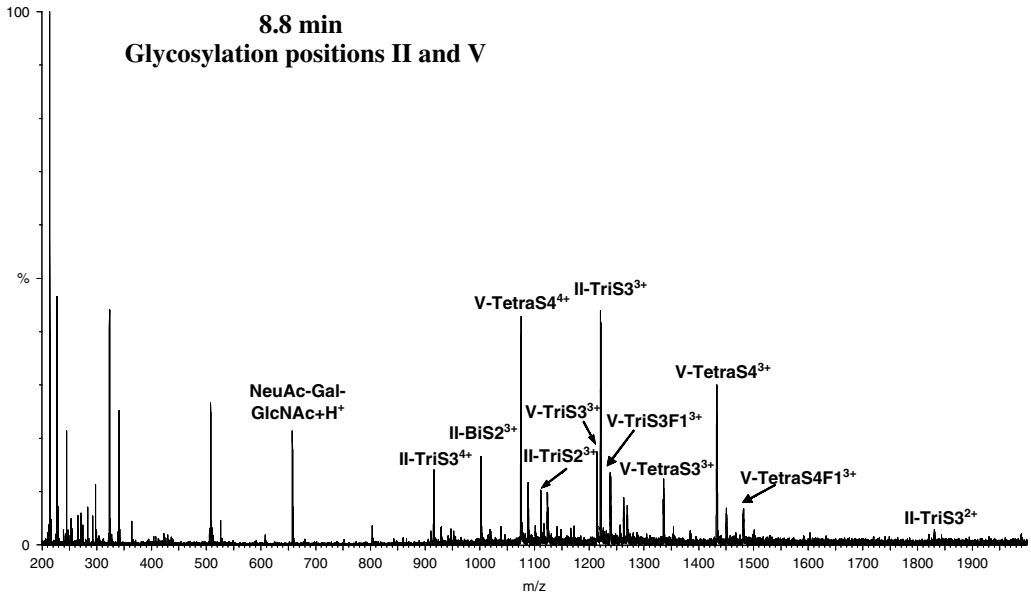

Figure 7. ESI-MS spectrum of a glycopeptide peak identified as site II and V from HPLC–ESI/MS analysis at retention time 8.8 min. The glycopeptid sequences were identified by mass matching and, later, MS/MS measurements.

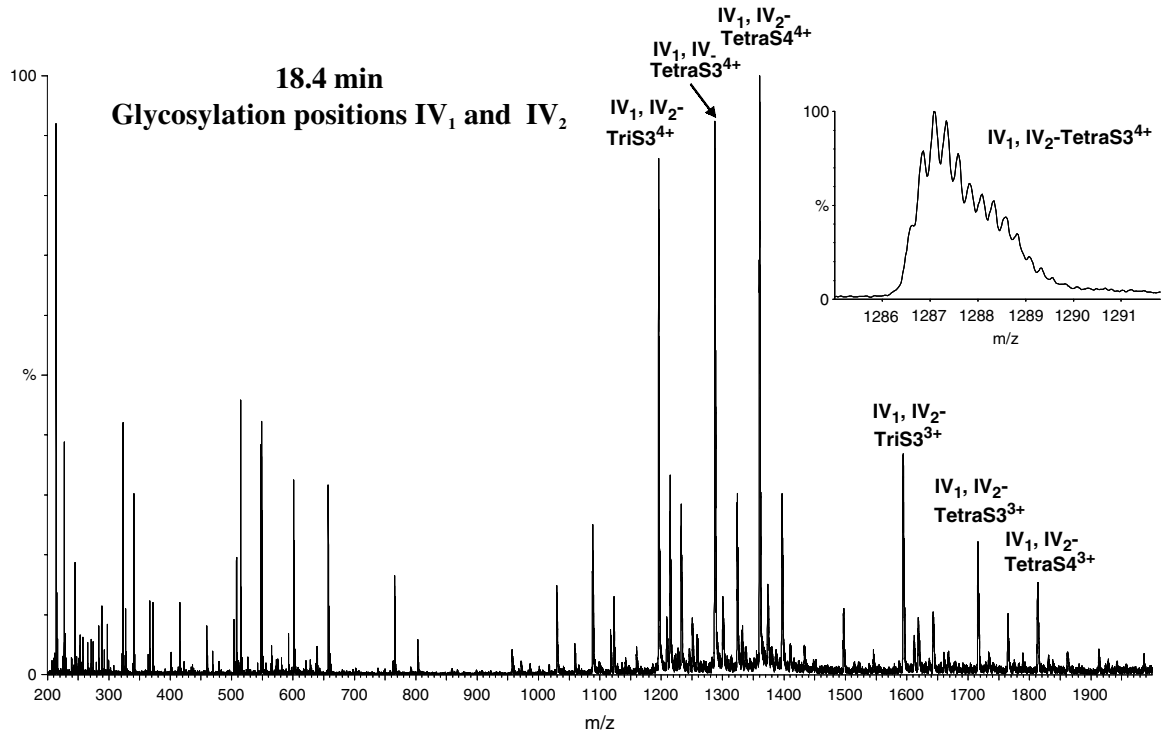

Figure 8. ESI-MS spectrum of a glycopeptide peak identified as site IV (IV_1 and IV_2) from HPLC–ESI/MS analysis at retention time 18.4 min. The glycopeptid sequences were identified by mass matching and MS/MS measurements.

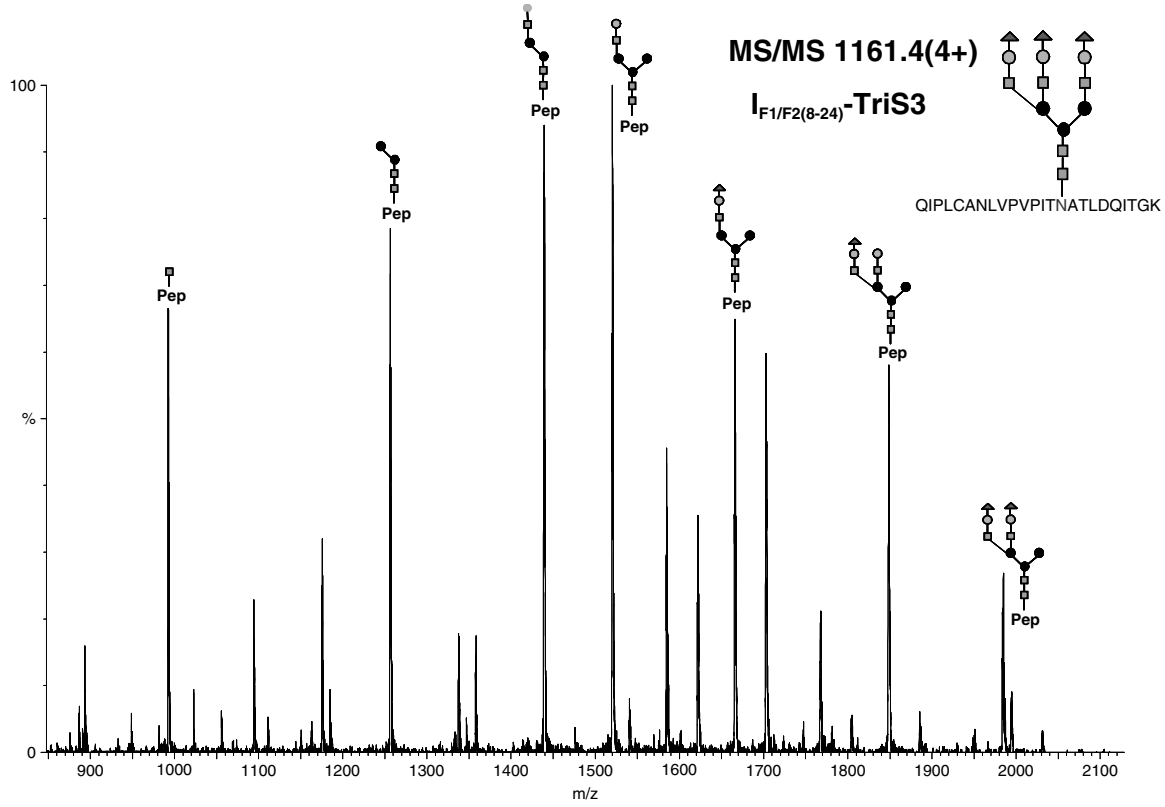

Figure 9. MS/MS spectra of glycopeptide from $I_{F1,F2(8-24)}$-TriS3, 1161.4 (4+) ion at 26.6 min. ◀ NeuAc, ○ Gal, □ GlcNAc, ● Man.

oligosaccharide composition. Glycopeptide-related peaks in the chromatogram have all been analyzed in a similar manner, and a large number of structures have been identified. The results are summarized in Table 1.

There are two nearly co-eluting glycopeptides (glycosylation sites II and V_1) (retention time 8.8 min), both of which have very short and similar peptide backbones. The full-scan electrospray spectrum of this chromatographic peak is

shown in Fig. 7. Glycosylation site II includes the peptide [34–39] with a mass of 795.3 Da; site V_1 is peptide [84–90] with a mass of 774.4 Da.

Another complexity is revealed by studying the spectrum of the peak at 18.4 min (Fig. 8). This peak contains glycosylation site IV, but the peptide backbone [69–83] contains two genetic variants (ORM1 and ORM2, identified as IV_1 and IV_2) with a mass difference of 4 Da. Most peaks in the spectrum are split by 4 Da and have an intensity ratio of 2 : 1. This suggests that genetic variants ORM1 : ORM2 are expressed in *ca* 2 : 1 ratio. This chromatographic peak has the highest intensity among glycopeptides observed, and the spectrum (Fig. 8) is dominated by triply and quadruply charged, tri- and tetra-antennary oligosaccharides with high sialic acid content.

Analysis of the AGP digest, discussed so far, successfully identified glycopeptides corresponding to glycosylation sites II, III, IV and V, but not those related to site I. On the other hand, the sugar characteristic ion chromatogram shown in Fig. 2(b) shows several other glycopeptides. Mass mapping, confirmed by subsequent MS/MS analysis (e.g. see spectrum in Fig. 9), indicates that most of these glycopeptides relate to site I glycosylation. Complexity in this case is derived from two factors. One, the known genetic variants express two different peptides related to site I glycosylation differing by four amino acid units (indicated either as [1–20] or [1–24] sequences, corresponding to ORM1∗S/ORM2∗A and ORM1∗F1/ORM1∗F2 alleles, respectively). Second, the present study shows that the AGP sample consists of a mixture of two polypeptide sequences, in which either four or seven of the *N*-terminal amino acids are missing. The origin and reason of the missing *N*-terminal amino acids were not studied (was outside the scope of the present work) but might be related to the commercial origin of the AGP sample. Site I glycosylation corresponds to the *N*-terminal tryptic peptide, so altogether four amino acid sequences are found corresponding to site I glycosylation, indicated as $I_{2(5-20)}$, $I_{1(5-24)}$, $I_{2(8-20)}$ and $I_{1(8-24)}$. All of these glycopeptides gave abundant peaks in the chromatogram (Fig. 2(b)). Glycosylation of all these variants were studied, showing very similar glycosylation patterns.

All major glycopeptide-containing chromatographic peaks were identified using this method and all five glycosylation sites have been found. The results are summarized in Table 1 and the glycosylation patterns corresponding to the five sites are shown in Fig. 10. Anomalies observed in digestion and in HPLC–MS studies of AGP are frequently encountered in studies of complex glycoproteins and were all reasonably explained: missed enzymatic cleavage site owing to neighboring bulky oligosaccharides, heterogeneity both of the peptide backbone and of the oligosaccharides and co-elution of various fragments.

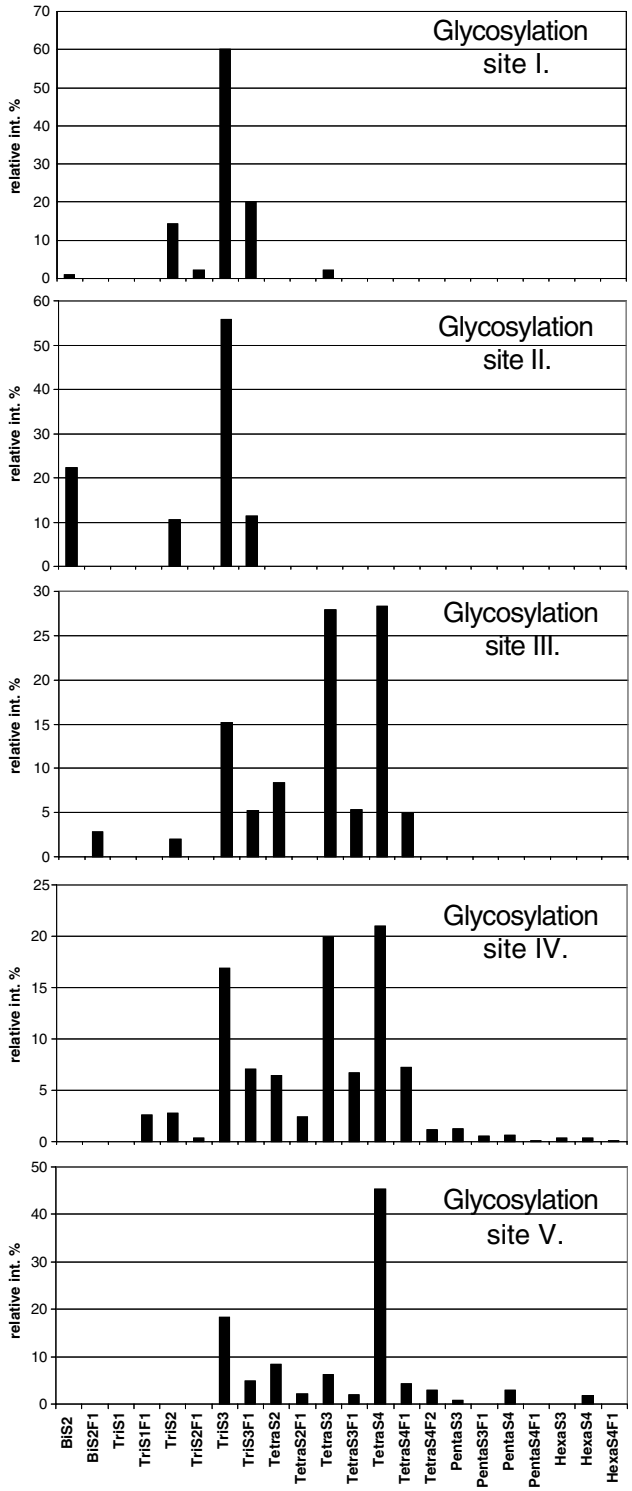

Figure 10. Relative intensities of glycans at the five different glycosylation sites of the AGP. In case of more than one possible glycopeptide in a glycosylation site, the relative intensities are averaged.

CONCLUSIONS

A new methodology has been developed for studying the site-specific glycosylation pattern of a complex glycoprotein, human AGP. Tryptic digestion is performed using a new anionic surfactant (RapiGest SF). This proved to be simple, efficient, well suited for subsequent analysis by mass spectrometry and can be recommended in the case of glycoproteins and other digestion-resistant proteins. Following digestion, the sample was directly injected onto a capillary HPLC–MS instrument. Glycopeptides were identified on the basis of their characteristic sugar oxonium ions (Fig. 2(b)). Retention time of glycopeptides is predominantly determined by the peptide chain and the

Table 1. Identified AGP glycopeptides from RapiGest-treated, trypsin-digested AGP. cam: carbamido-methyl group

	Peptide residues	Oligosaccharides	Observed mass	Calculated mass
Glycosylation site I	ORM1∗S/ORM2∗A [8–20]	BiS2	1205.61 (3+)/1807.07 (2+)	1205.26 (3+)/1807.39 (2+)
19.5 min	$M = 1408.66$	TriS1	1230.28 (3+)	1230.28 (3+)
		TriS2	995.74 (4+)/1327.35 (3+)	995.48 (4+)/1326.97 (3+)
		TriS2F1	1032.26 (4+)/1376.07 (3+)	1031.99 (4+)/1375.66 (3+)
		TriS3	1068.54 (4+)/1424.41 (3+)	1068.25 (4+)/1424.00 (3+)
		TriS3F1	1105.01 (4+)/1473.11 (3+)	1104.76 (4+)/1472.69 (3+)
		TetraS3	1160.10 (4+)/1546.39 (3+)	1159.53 (4+)/1545.71 (3+)
21.3 min	ORM1∗S/ORM2∗A [5–20] cam	TriS2	1082.11 (4+)/1442.41 (3+)	1081.72 (4+)/1141.97 (3+)
	$M = 1752.98$	TriS3	1154.90 (4+)/1539.49 (3+)	1154.50 (4+)/1539.00 (3+)
		TriS3F1	1191.42 (4+)/1588.24 (3+)	1191.01 (4+)/1587.89 (3+)
26.6 min	ORM1∗F1/ORM1∗F2 [8–24]	BiS2	997.30 (4+)/1329.36 (3+)	997.19 (4+)/1329.26 (3+)
	$M = 1778.0$	TriS2	1088.60 (4+)/1451.22 (3+)	1088.47 (4+)/1450.97 (3+)
		TriS2F1	1125.11 (4+)/1499.90 (3+)	1124.99 (4+)/1499.66 (3+)
		TriS3	1161.42 (4+)/1548 (3+)	1161.25 (4+)/1548.00 (3+)
		TriS3F1	1197.96 (4+)/1596.84 (3+)	1197.76 (4+)/1596.68 (3+)
		TetraS3	1252.74 (4+)/1669.85 (3+)	1252.53 (4+)/1669.71 (3+)
28.7 min	ORM1∗F1/ORM1∗F2 [5–24] cam	TriS2	1174.95 (4+)/1566.16 (3+)	1174.58 (4+)/1565.77 (3+)
	$M = 2124.41$	TriS3	1247.71 (4+)/1663.3 (3+)	1247.35 (4+)/1662.80 (3+)
		TriS3F1	1284.22 (4+)/1712.01 (4+)	1283.87 (4+)/1711.49 (3+)
Glycosylation site II	[34–39]	BiS2	1001.36 (3+)	1001.04 (3+)
8.8 min	$M = 795.3$	TriS2	1123.12 (3+)	1122.75 (3+)
		TriS3	1220.16 (3+)/915.32 (4+)	1219.78 (3+)/915.088 (4+)
		TriS3F1	1268.88 (3+)/951.83 (4+)	1268.47 (3+)/951.602 (4+)
Glycosylation site III	[40–63] (1 missed cleavage)	BiS2F1	1312.98 (4+)	1312.32 (4+)
42.8 min	$M = 2894.44$	TriS2	1367.69 (4+)	1367.09 (4+)
		TriS3	1440.44 (4+)/1152.63 (5+)	1439.86 (4+)/1152.09 (5+)
		TriS3F1	1476.97 (4+)	1476.38 (4+)
		TetraS2	1458.94 (4+)	1458.37 (4+)
		TetraS3	1531.66 (4+)/1225.72 (5+)	1534.14 (4+)/1225.11 (5+)
		TetraS3F1	1568.40 (4+)	1567.66 (4+)
		TetraS4	1604.5 (4+)/1283.99 (5+)	1603.92 (4+)/1283.34 (5+)
		TetraS4F1	1641.15 (4+)	1640.43 (4+)
Glycosylation site IV	ORM1 [69–83] cam	TriS2	1122.83 (4+)/1496.44 (3+)	1122.20 (4+)/1495.93 (3+)
18.4 min	$M = 1914.89$	TriS2F1	1159.19 (4+)	1158.71 (4+)
		TriS3	1195.61 (4+)/1593.45 (3+)	1194.97 (4+)/1592.96 (3+)
		TriS3F1	1231.92 (4+)/1642.24 (3+)	1231.49 (4+)/1641.65 (3+)
		TetraS2	1213.88 (4+)/1618.2 (3+)	1213.48 (4+)/1617.63 (3+)
		TetraS2F1	1250.48 (4+)/1666.75 (3+)	1249.99 (4+)/1666.32 (3+)
		TetraS3	1286.60 (4+)/1715.44 (3+)	1286.26 (4+)/1714.67 (3+)
		TetraS3F1	1323.24 (4+)/1764.04 (3+)	1322.77 (4+)/1763.36 (3+)
		TetraS4	1359.74 (4+)/1812.63 (3+)	1359.03 (4+)/1811.71 (3+)
		TetraS4F1	1395.93 (4+)/1861.11 (3+)	1395.54 (4+)/1860.39 (3+)
		TetraS4F2	1432.55 (4+)	1432.06 (4+)
		PentaS3	1377.94 (4+)/1836.74 (3+)	1377.54 (4+)/1836.38 (3+)
		PentaS3F1	1414.53 (4+)	1414.05 (4+)
		PentaS4	1451.01 (4+)/1934.38 (3+)	1450.31 (4+)/1933.42 (3+)
		PentaS4F1	1487.59 (4+)	1486.83 (4+)
		HexaS3	1469.22 (4+)	1468.82 (4+)
		HexaS4	1542.10 (4+)	1541.39 (4+)
		HexaS4F1	1578.50 (4+)	1578.11 (4+)
18.4 min	ORM 2 [69–83] cam	TriS1F1	1087.86 (4+)	1087.18 (4+)
	$M = 1919.86$	TriS2	1123.82 (4+)/1498.00 (3+)	1123.44 (4+)/1497.59 (3+)
		TriS3	1196.60 (4+)/1595.19 (3+)	1196.21 (4+)/1594.62 (3+)

Table 1. (*Continued*)

	Peptide residues	Oligosaccharides	Observed mass	Calculated mass
		TriS3F1	1233.12 (4+)/1644.02 (3+)	1232.73 (4+)/1643.31 (3+)
		TetraS2	1215.14 (4+)	1214.72 (4+)
		TetraS2F1	1251.87 (4+)	1251.24 (4+)
		TetraS3	1287.88 (4+)/1716.95 (3+)	1287.50 (4+)/1716.33 (3+)
		TetraS3F1	1324.45 (4+)/1765.75 (3+)	1324.01 (4+)/1765.02 (3+)
		TetraS4	1360.71 (4+)/1813.91 (3+)	1360.27 (4+)/1813.36 (3+)
		TetraS4F1	1397.47 (4+)	1396.79 (4+)
		TetraS4F2	1434.10 (4+)	1433.30 (4+)
		PentaS3	1379.46 (4+)	1378.78 (4+)
		PentaS3F1	1416.09 (4+)	1415.29 (4+)
		PentaS4	1452.03 (4+)	1451.56 (4+)
		HexaS3	1471.01 (4+)	1470.06 (4+)
		HexaS4	1543.86 (4+)	1543.09 (4+)
Glycosylation site V	ORM1 [84–90]	TriS3	1213.55 (3+)/910.37 (4+)	1212.79 (3+)/909.84 (4+)
8.8 min	$M = 774.36$	TriS3F1	1262.21 (3+)	1261.47 (3+)
		TetraS2	1237.79 (3+)	1237.46 (3+)
		TetraS2F1	1286.53 (3+)	1285.15 (3+)
		TetraS3	1335.27 (3+)	1334.50 (3+)
		TetraS3F1	1383.89 (3+)	1383.18 (3+)
		TetraS4	1074.52 (4+)/1432.37 (3+)	1073.90 (4+)/1431.53 (3+)
		TetraS4F1	1111.03 (4+)/1481.10 (3+)	1110.41 (3+)/1480.22 (3+)
		TetraS4F2	1147.57 (4+)/1529.73 (3+)	1146.92 (4+)/1528.90 (3+)
		PentaS3	1456.85 (3+)	1456.21 (3+)
		PentaS4	1165.82 (4+)/1554.17 (3+)	1165.19 (4+)/1553.24 (3+)
		HexaS4	1257.08 (4+)	1256.46 (4+)

presence of various glycans results in peak broadening. Glycan structures were identified on the basis of molecular mass distribution (using low-energy single-stage mass spectra of various chromatographic peaks), followed by confirmation using MS/MS.

Using the procedure discussed above, all glycosylation sites of AGP have been characterized, and altogether 80 different glycopeptides were identified, about three times as many as known previously[13,14,22,27] (Table 1). Relative intensity of the identified glycopeptides can be regarded as a semiquantitative measure for the site distribution of glycosylation, and this is shown in Fig. 10. Glycopeptide variants $I_{2(5-20)}$, $I_{1(5-24)}$, $I_{2(8-20)}$, $I_{1(8-24)}$ and IV_1, IV_2, respectively, showed virtually identical glycosylation patterns; Fig. 10 shows the average results for site I and IV. The results are in good qualitative agreement with previous findings,[13,22,27] indicating lower branching for sites I and II and more highly branched oligosaccharides for sites III, IV and V. The present results show a fairly similar glycosylation pattern for sites I and II, the triantennary oligosaccharide TriS3 being most abundant. Sites III and IV show a markedly different pattern, where tetra-antennary structures prevail (TetraS3 and TetraS4). In the case of site V, the tetra-antennary structure TetraS4 is predominant. Appearance of even more highly branched structures (or those with elongated antennae containing additional GlcNAc-Gal units) is a remarkable feature of glycosylation at sites IV and V.

The marked site specificity of glycosylation may be related to the 3D structure and biosynthesis of AGP (e.g. limited access of glycosylating enzymes to various sites). Site-specific glycosylation may affect the biological function of AGP and reflect the pathophysiological state of the organism, which is planned to be explored in the near future.

Acknowledgements

Authors are grateful to Pál T. Szabó (Chemical Research Center, Budapest, Hungary) for the helpful discussions. The work was sponsored by the Center of Excellence Project (CBCH, No. QLK2-CT-2002-90436), by the Education Ministry of Hungary NKFP MediChem2 1/A/005/2004 and by the Italian-Hungarian Joint Project 2004/2006 (MTA707).

REFERENCES

1. Dwek RA. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* 1996; **96**: 683.
2. Rudd PM, Dwek RA. Glycosylation: Heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* 1997; **32**: 1.
3. Durand G, Seta N. Protein glycosylation and diseases: Blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clin. Chem.* 2000; **46**: 795.
4. Lijima S, Shiba K, Kimura M, Nagai K, Iwai T. Changes of alpha(1)-acid glycoprotein microheterogeneity in acute inflammation stages analyzed by isoelectric focusing using serum obtained postoperatively. *Electrophoresis* 2000; **21**: 753.
5. Spellman MW. Carbohydrate characterization of recombinant glycoproteins of pharmaceutical interest. *Anal. Chem.* 1990; **62**: 1714.
6. Harvey DJ. Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics.* 2001; **1**: 311.
7. Harvey DJ. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom. Rev.* 1999; **18**: 349.

8. Harvey DJ. Collision-induced fragmentation of underivatized N-linked carbohydrates ionized by electrospray. *J. Mass Spectrom.* 2000; **35**: 1178.
9. Harvey DJ. Electrospray mass spectrometry and fragmentation of N-linked carbohydrates derivatized at the reducing terminus. *J. Am. Soc. Mass Spectrom.* 2000; **11**: 900.
10. Harvey DJ. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *Int. J. Mass spectrom.* 2003; **226**: 1.
11. Conboy JJ, Henion JD. The determination of glycopeptides by liquid-chromatography mass-spectrometry with collision-induced dissociation. *J. Am. Soc. Mass Spectrom.* 1992; **3**: 804.
12. Huddleston MJ, Bean MF, Carr SA. Collisional fragmentation of glycopeptides by electrospray ionization LCMS and LCMS/MS – methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* 1993; **65**: 877.
13. Juhasz P, Martin SA. The utility of nonspecific proteases in the characterization of glycoproteins by high-resolution time-of-flight mass spectrometry. *Int. J. Mass spectrom.* 1997; **169**: 217.
14. Dage JL, Ackermann BL, Halsall HB. Site localization of sialyl Lewis(x) antigen on alpha(1)-acid glycoprotein by high performance liquid chromatography electrospray mass spectrometry. *Glycobiology* 1998; **8**: 755.
15. Park ZY, Russell DH. Thermal denaturation: A useful technique in peptide mass mapping. *Anal. Chem.* 2000; **72**: 2667.
16. Park ZY, Russell DH. Identification of individual proteins in complex protein mixtures by high-resolution, high-mass-accuracy MALDI TOF-mass spectrometry analysis of in-solution thermal denaturation/enzymatic digestion. *Anal. Chem.* 2001; **73**: 2558.
17. Russell WK, Park ZY, Russell DH. Proteolysis in mixed organic-aqueous solvent systems: Applications for peptide mass mapping using mass spectrometry. *Anal. Chem.* 2001; **73**: 2682.
18. Stover T, Amari JV, Mazsaroff I, Gilar M, Yu YQ, Gebler JC. Novel characterization tool for Mab digestion – Technical note: RapiGest SF denaturant tool for improved trypsin digestion of monoclonal antibodies. *Genet. Eng. News* 2003; **23**: 48.
19. Yu YQ, Gilar M, Lee PJ, Bouvier ESP, Gebler JC. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* 2003; **75**: 6023.
20. Nomura E, Katsuta K, Ueda T, Toriyama M, Mori T, Inagaki N. Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping. *J. Mass Spectrom.* 2004; **39**: 202.
21. Schmid K, Baumann P, Müller WE, Eap CB, Tillement JP (eds). In *Alpha1-acid Glycoprotein. Genetics, Biochemistry, Physiological Functions, and Pharmacology*. Alain R. Liss Incorporated: New York, 1989; 7.
22. Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. *Biochim. Biophys. Acta – Protein Struct. Mol. Enzymol.* 2000; **1482**: 157.
23. Dente L, Ruther U, Tripodi M, Wagner EF, Cortese R. Expression of human alpha-1-acid glycoprotein genes in cultured-cells and in transgenic mice. *Genes Dev.* 1988; **2**: 259.
24. Hashimoto S, Asao T, Takahashi J, Yagihashi Y, Nishimura T, Saniabadi AR, Poland DCW, van Dijk W, Kuwano H, Kochibe N, Yazawa S. Alpha(1)-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis. *Cancer* 2004; **101**: 2825.
25. Kremmer T, Szollosi E, Boldizsar M, Vincze B, Ludanyi K, Imre T, Schlosser G, Vekey K. Liquid chromatographic and mass spectrometric analysis of human serum acid alpha-1-glycoprotein. *Biomed. Chromatogr.* 2004; **18**: 323.
26. Duche JC, Urien S, Simon N, Malaurie E, Monnet I, Barre J. Expression of the genetic variants of human alpha-1-acid glycoprotein in cancer. *Clin. Biochem.* 2000; **33**: 197.
27. Treuheit MJ, Costello CE, Halsall HB. Analysis of the five glycosylation sites of human alpha-1-acid glycoprotein. *Biochem. J.* 1992; **283**: 105.
28. Jiang H, Desaire H, Butnev VY, Bousfield GR. Glycoprotein profiling by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2004; **15**: 750.
29. Ritchie MA, Gill AC, Deery MJ, Lilley K. Precursor ion scanning for detection and structural characterization of heterogeneous glycopeptide mixtures. *J. Am. Soc. Mass Spectrom.* 2002; **13**: 1065.